Mixing times via super-fast coupling

Yevgeniy Kovchegov

Department of Mathematics
Oregon State University

Outline

Outline

Coauthor

Paper.

This presentation is based on joint work with R.Burton.

History

Diaconis and Shahshahani (early 80's)

The mixing time for shuffling a deck of n cards by random transpositions is of order $O(n \log(n))$ with cut-off asymptotics at $\frac{1}{2} n \log(n)$.

Method used: relatively rarified mathematical residential district of representation theory.

The Problem

Open problem (Y.Peres)

Provide a coupling proof of $O(n\log(n))$ mixing rate.

Continuous time: $<\boxed{\mathbf{a}},\boxed{\mathbf{b}}>$ has rate $\frac{2}{n^2}$, i.e. $<\boxed{\mathbf{a}},\boxed{\mathbf{b}}>$ and $<\boxed{\mathbf{b}},\boxed{\mathbf{a}}>$ happen with rate $\frac{1}{n^2}$ each.
Transposition $<\boxed{\mathbf{a}},\boxed{\mathbf{a}}>$ has rate $\frac{1}{n^2}$.

Outline

Definition of Mixing Time

Total variation distance:

$$\|\mu - \nu\|_{TV} := \frac{1}{2} \sum_x |\mu(x) - \nu(x)|$$

Mixing time: process $\{X_t\}$ with stationary distribution π. If $X_t \sim \nu_t = \nu_0 P_t$,

$$t_{mix} := \inf \left\{ t \;:\; \|\nu_t - \pi\|_{TV} \leq \frac{1}{4}, \quad \text{all } \nu_0 \right\}$$

Mixing time via coupling

$$X_t \sim \nu_0 P_t \text{ and } Y_t \sim \mu_0 P_t.$$

If $T_{coupling}$ is the coupling time for (X_t, Y_t), then

$$\|\nu_0 P_t - \mu_0 P_t\|_{TV} \leq P[T_{coupling} > t]$$

Goal: bound $P[T_{coupling} > t]$ for all μ_0.
Here $\boxed{\max_x \|\nu_0 P_t - x P_t\|_{TV} \geq \|\nu_0 P_t - \pi\|_{TV}}$ as

$$\sum_x \pi(x) \|\nu_0 P_t - x P_t\|_{TV} \geq \|\nu_0 P_t - \pi\|_{TV}$$

by convexity of $f(x) = \|\nu_0 P_t - x P_t\|_{TV}$.

Notations

Notations

$< \cdot, \cdot >$	- transpositions
$< \boxed{\textbf{a}}, \cdot >$	- transposition initiated by card $\boxed{\textbf{a}}$
$\begin{pmatrix} A_t \\ B_t \end{pmatrix}$	- the coupled process
$\ll \cdot, \cdot \gg$	- transpositions in $\begin{pmatrix} A_t \\ B_t \end{pmatrix}$

Coupling #1. (Aldous and Fill) $\ll \boxed{\mathbf{a}}, i \gg$: moves card $\boxed{\mathbf{a}}$ to location i in both processes, A_t and B_t.
Mixing: order $O(n^2)$ instead of $O(n \log n)$;

$$E[T_{coupling}] \approx \sum_{d=2}^{n} \frac{n^2}{d^2} \approx \left(\frac{\pi^2}{6} - 1\right) n^2$$

Problem: slows down significantly when the number of discrepancies is small enough,

Coupling #2. (equivalent to prev.):
- applying transposition $\ll \boxed{\mathbf{a}}, i \gg$ *if* $\boxed{\mathbf{a}}$ *is not coupled*,
- applying transposition $\ll \boxed{\mathbf{a}}, \boxed{\mathbf{b}} \gg$ *if* $\boxed{\mathbf{a}}$ *is coupled*

Coupling #2 can be improved to match $O(n \log n)$
Transpositions $\ll \boxed{\mathbf{a}}, \boxed{\mathbf{b}} \gg$ called **label-to-label**.

Group invariance

X_m is a Markov Chain on a discrete group G. The chain is **group invariant** if

$$dist(\gamma X_{m+1} | X_m = \alpha) = dist(X_{m+1} | X_m = \gamma^{-1}\alpha)$$

for all $\gamma, \alpha \in S_n$.

In other words,

label-to-label $\ll \boxed{\mathbf{a}}, \boxed{\mathbf{b}} \gg$ can be ignored.

Reason: don't change cycle structure.
The situation is invariant under label-to-label transpositions.

If label-to-label $\ll \boxed{\mathbf{a}}, \boxed{\mathbf{b}} \gg$ is applied before label-to-location $\ll \boxed{\mathbf{b}}, i \gg$. We do not relabel, and do not reselect i.

Vocabulary.

association map - hidden association between positions/locations in the top process and positions/locations in the bottom process that will be used to establish the rates for the coupled process

Two discrepancies ($d = 2$) at d_1 and d_2:

$$\begin{array}{lcccccccccc} A_t: & \dots & \boxed{4} & \boxed{6} & \boxed{\mathbf{b}} & \boxed{9} & \boxed{\mathbf{a}} & \boxed{8} & \boxed{\mathbf{a1}} & \boxed{2} & \dots \\ B_t: & \dots & \boxed{4} & \boxed{6} & \boxed{\mathbf{a}} & \boxed{9} & \boxed{\mathbf{b}} & \boxed{8} & \boxed{\mathbf{a1}} & \boxed{2} & \dots \\ & & & & \uparrow & & \uparrow & & \uparrow & & \\ & & & & d_2 & & d_1 & & i_1 & & \end{array}$$

Label-to-location coupling:

$$E[T_{coupling}] = \frac{n^2}{4} \quad \text{– too large.}$$

Jump $\ll \boxed{\mathbf{a}}, i_1 \gg$ of $\boxed{\mathbf{a}}$ to random location i_1 at exponential time t_1:

From

$$\begin{array}{lcccccccccc}
A_t: & \dots & \boxed{4} & \boxed{6} & \boxed{\mathbf{b}} & \boxed{9} & \boxed{\mathbf{a}} & \boxed{8} & \boxed{\mathbf{a1}} & \boxed{2} & \dots \\
B_t: & \dots & \boxed{4} & \boxed{6} & \boxed{\mathbf{a}} & \boxed{9} & \boxed{\mathbf{b}} & \boxed{8} & \boxed{\mathbf{a1}} & \boxed{2} & \dots \\
 & & & & \uparrow & & \uparrow & & \uparrow & & \\
 & & & & d_2 & & d_1 & & i_1 & &
\end{array}$$

to

$$\begin{array}{lcccccccccc}
A_t: & \dots & \boxed{4} & \boxed{6} & \boxed{\mathbf{b}} & \boxed{9} & \boxed{\mathbf{a1}} & \boxed{8} & \boxed{\mathbf{a}} & \boxed{2} & \dots \\
B_t: & \dots & \boxed{4} & \boxed{6} & \boxed{\mathbf{a1}} & \boxed{9} & \boxed{\mathbf{b}} & \boxed{8} & \boxed{\mathbf{a}} & \boxed{2} & \dots \\
 & & & & \uparrow & & \uparrow & & \uparrow & & \\
 & & & & d_2 & & d_1 & & i_1 & &
\end{array}$$

Different way of saying the same:

Start with

$$
\begin{array}{lccccccccccc}
A_t: & \dots & \boxed{4} & \boxed{6} & \boxed{\mathbf{b}} & \boxed{9} & \boxed{\mathbf{a1}} & \boxed{8} & \boxed{\mathbf{a}} & \boxed{2} & \dots \\
B_t: & \dots & \boxed{4} & \boxed{6} & \boxed{\mathbf{a1}} & \boxed{9} & \boxed{\mathbf{b}} & \boxed{8} & \boxed{\mathbf{a}} & \boxed{2} & \dots \\
 & & & & \uparrow & & \uparrow & & \uparrow & & \\
 & & & & d_2/i_1 & & i_1/d_1 & & d_1/d_2 & &
\end{array}
$$

where at time t_1 the locations relabel according to

$$
\boxed{\begin{array}{l}
d_1/d_2 \longrightarrow i_1 \\
i_1/d_1 \longrightarrow d_1 \\
d_2/i_1 \longrightarrow d_2
\end{array}}.
$$

Different way of saying the same:

Jump $\ll \boxed{\mathbf{a}}, i_1 \gg$ at time $t_1 \sim$ **exponential** $\left(\frac{1}{n}\right)$.
From

$$
\begin{array}{llcccccccccc}
A_t : & \dots & \boxed{4} & \boxed{6} & \boxed{\mathbf{b}} & \boxed{9} & \boxed{\mathbf{a1}} & \boxed{8} & \boxed{\mathbf{a}} & \boxed{2} & \dots \\
B_t : & \dots & \boxed{4} & \boxed{6} & \boxed{\mathbf{a1}} & \boxed{9} & \boxed{\mathbf{b}} & \boxed{8} & \boxed{\mathbf{a}} & \boxed{2} & \dots \\
 & & & & \uparrow & & \uparrow & & \uparrow & & \\
 & & & & d_2/i_1 & & i_1/d_1 & & d_1/d_2 & &
\end{array}
$$

to

$$
\begin{array}{llcccccccccc}
A_t : & \dots & \boxed{4} & \boxed{6} & \boxed{\mathbf{b}} & \boxed{9} & \boxed{\mathbf{a1}} & \boxed{8} & \boxed{\mathbf{a}} & \boxed{2} & \dots \\
B_t : & \dots & \boxed{4} & \boxed{6} & \boxed{\mathbf{a1}} & \boxed{9} & \boxed{\mathbf{b}} & \boxed{8} & \boxed{\mathbf{a}} & \boxed{2} & \dots \\
 & & & & \uparrow & & \uparrow & & \uparrow & & \\
 & & & & d_2 & & d_1 & & i_1 & &
\end{array}
$$

The Association Map.

The following **association map** will determine jumps of $\boxed{\textbf{a1}}$.

$$
\begin{array}{lccccccccccc}
A_t : & \ldots & \boxed{4} & \boxed{6} & \boxed{\textbf{b}} & \boxed{9} & \boxed{\textbf{a1}} & \boxed{8} & \boxed{\textbf{a}} & \boxed{2} & \ldots \\
B_t : & \ldots & \boxed{4} & \boxed{6} & \boxed{\textbf{a1}} & \boxed{9} & \boxed{\textbf{b}} & \boxed{8} & \boxed{\textbf{a}} & \boxed{2} & \ldots
\end{array}
$$

Card $\boxed{\textbf{a1}}$ will jump to position i_2 on the assoc. map at time t_2, even if $t_2 < t_1$.

$$\begin{array}{llcccccccccl} A_t: & \dots & \boxed{4} & \boxed{6} & \boxed{\mathbf{b}} & \boxed{9} & \boxed{\mathbf{a1}} & \boxed{8} & \boxed{\mathbf{a}} & \boxed{2} & \dots \\ B_t: & \dots & \boxed{4} & \boxed{6} & \boxed{\mathbf{a1}} & \boxed{9} & \boxed{\mathbf{b}} & \boxed{8} & \boxed{\mathbf{a}} & \boxed{2} & \dots \\ & & & & \uparrow & & \uparrow & & \uparrow & & \\ & & & & d_2^* & & d_1^* & & i_1^* & & \end{array}$$

Now $i_2 \neq i_1^*$ and

$$t_2 \sim \textbf{exponential}\left((1 - 1/n) \cdot \frac{1}{n}\right)$$

$\ll \boxed{\mathbf{a1}}, i_1^* \gg = \ll \boxed{\mathbf{a1}}, \boxed{\mathbf{a}} \gg$ is label-to-label, we can skip.

If $i_1 = d_1$ or d_2, discrepancies cancel at t_1;
if $i_2^* = d_1^*$ or d_2^*, discrepancies cancel on the assoc. map at t_2.
If $t_1 < t_2$, $\boxed{\text{assoc. map} \to \text{real picture}}$ at t_1, we create one more assoc. map.

Case $t_2 < t_1$, and $i_2^* = d_2^*$. On association map:

Start with

$$
\begin{array}{lcccccccccc}
A_t: & \dots & \boxed{4} & \boxed{6} & \boxed{\mathbf{b}} & \boxed{9} & \boxed{\mathbf{a1}} & \boxed{8} & \boxed{\mathbf{a}} & \boxed{2} & \dots \\
B_t: & \dots & \boxed{4} & \boxed{6} & \boxed{\mathbf{a1}} & \boxed{9} & \boxed{\mathbf{b}} & \boxed{8} & \boxed{\mathbf{a}} & \boxed{2} & \dots \\
 & & & & \uparrow & & \uparrow & & \uparrow & & \\
 & & & & d_2/i_1 & & i_1/d_1 & & d_1/d_2 & &
\end{array}
$$

At time t_2:

$$
\begin{array}{lcccccccccc}
A_t: & \dots & \boxed{4} & \boxed{6} & \boxed{\mathbf{a1}} & \boxed{9} & \boxed{\mathbf{b}} & \boxed{8} & \boxed{\mathbf{a}} & \boxed{2} & \dots \\
B_t: & \dots & \boxed{4} & \boxed{6} & \boxed{\mathbf{a1}} & \boxed{9} & \boxed{\mathbf{b}} & \boxed{8} & \boxed{\mathbf{a}} & \boxed{2} & \dots \\
 & & & & \uparrow & & \uparrow & & \uparrow & & \\
 & & & & d_2/i_1 & & i_1/d_1 & & d_1/d_2 & &
\end{array}
$$

Case $t_2 < t_1$, and $i_2^* = d_2^*$. On association map:

At time t_2:

$$
\begin{array}{lcccccccccc}
A_t: & \dots & \boxed{4} & \boxed{6} & \boxed{\mathbf{a1}} & \boxed{9} & \boxed{\mathbf{b}} & \boxed{8} & \boxed{\mathbf{a}} & \boxed{2} & \dots \\
B_t: & \dots & \boxed{4} & \boxed{6} & \boxed{\mathbf{a1}} & \boxed{9} & \boxed{\mathbf{b}} & \boxed{8} & \boxed{\mathbf{a}} & \boxed{2} & \dots \\
 & & & & \uparrow & & \uparrow & & \uparrow & & \\
 & & & & d_2/i_1 & & i_1/d_1 & & d_1/d_2 & &
\end{array}
$$

At time t_1:

$$
\begin{array}{lcccccccccc}
A_t: & \dots & \boxed{4} & \boxed{6} & \boxed{\mathbf{a1}} & \boxed{9} & \boxed{\mathbf{b}} & \boxed{8} & \boxed{\mathbf{a}} & \boxed{2} & \dots \\
B_t: & \dots & \boxed{4} & \boxed{6} & \boxed{\mathbf{a1}} & \boxed{9} & \boxed{\mathbf{b}} & \boxed{8} & \boxed{\mathbf{a}} & \boxed{2} & \dots \\
 & & & & \uparrow & & \uparrow & & \uparrow & & \\
 & & & & d_2 & & d_1 & & i_1 & &
\end{array}
$$

Case $t_2 < t_1$, and $i_2^* = d_2^*$. Same evolution, original association:

Start with

$$
\begin{array}{lccccccccccc}
A_t: & \dots & \boxed{4} & \boxed{6} & \boxed{\mathbf{b}} & \boxed{9} & \boxed{\mathbf{a}} & \boxed{8} & \boxed{\mathbf{a1}} & \boxed{2} & \dots \\
B_t: & \dots & \boxed{4} & \boxed{6} & \boxed{\mathbf{a}} & \boxed{9} & \boxed{\mathbf{b}} & \boxed{8} & \boxed{\mathbf{a1}} & \boxed{2} & \dots \\
 & & & & \uparrow & & \uparrow & & \uparrow & & \\
 & & & & d_2 & & d_1 & & i_1 & &
\end{array}
$$

At time t_2:

$$
\begin{array}{lccccccccccc}
A_t: & \dots & \boxed{4} & \boxed{6} & \boxed{\mathbf{a1}} & \boxed{9} & \boxed{\mathbf{a}} & \boxed{8} & \boxed{\mathbf{b}} & \boxed{2} & \dots \\
B_t: & \dots & \boxed{4} & \boxed{6} & \boxed{\mathbf{a}} & \boxed{9} & \boxed{\mathbf{b}} & \boxed{8} & \boxed{\mathbf{a1}} & \boxed{2} & \dots \\
 & & & & \uparrow & & \uparrow & & \uparrow & & \\
 & & & & d_2 & & d_1 & & i_1 & &
\end{array}
$$

Case $t_2 < t_1$, and $i_2^* = d_2^*$. Same evolution, original association:

At time t_2:

$$
\begin{array}{lcccccccccc}
A_t: & \ldots & \boxed{4} & \boxed{6} & \boxed{\mathbf{a1}} & \boxed{9} & \boxed{\mathbf{a}} & \boxed{8} & \boxed{\mathbf{b}} & \boxed{2} & \ldots \\
B_t: & \ldots & \boxed{4} & \boxed{6} & \boxed{\mathbf{a}} & \boxed{9} & \boxed{\mathbf{b}} & \boxed{8} & \boxed{\mathbf{a1}} & \boxed{2} & \ldots \\
 & & & & \uparrow & & \uparrow & & \uparrow & & \\
 & & & & d_2 & & d_1 & & i_1 & &
\end{array}
$$

At time t_1:

$$
\begin{array}{lcccccccccc}
A_t: & \ldots & \boxed{4} & \boxed{6} & \boxed{\mathbf{a1}} & \boxed{9} & \boxed{\mathbf{b}} & \boxed{8} & \boxed{\mathbf{a}} & \boxed{2} & \ldots \\
B_t: & \ldots & \boxed{4} & \boxed{6} & \boxed{\mathbf{a1}} & \boxed{9} & \boxed{\mathbf{b}} & \boxed{8} & \boxed{\mathbf{a}} & \boxed{2} & \ldots \\
 & & & & \uparrow & & \uparrow & & \uparrow & & \\
 & & & & d_2 & & d_1 & & i_1 & &
\end{array}
$$

Faster coupling.

The new coupling time for two discrepancies:

$$E[T_{coupling}] \approx \frac{n^2}{8}$$

Chain of association maps.

$\ll \boxed{\mathbf{a1}}, i_2 \gg$ occurs at t_2

$$
\begin{array}{lccccccccccc}
A_t: & \dots & \boxed{\mathbf{a2}} & \boxed{6} & \boxed{\mathbf{b}} & \boxed{9} & \boxed{\mathbf{a1}} & \boxed{8} & \boxed{\mathbf{a}} & \boxed{2} & \dots \\
B_t: & \dots & \boxed{\mathbf{a2}} & \boxed{6} & \boxed{\mathbf{a1}} & \boxed{9} & \boxed{\mathbf{b}} & \boxed{8} & \boxed{\mathbf{a}} & \boxed{2} & \dots \\
 & & \uparrow & & \uparrow & & \uparrow & & \uparrow & & \\
 & & i_2 & & d_2^* & & d_1^* & & i_1^* & &
\end{array}
$$

d_1^* is i_1/d_1 before t_1, and d_1 after t_1;
d_2^* is d_2/i_1 before t_1, and d_2 after t_1;
i_1^* is d_1/d_2 before t_1, and i_1 after t_1.

Chain of association maps.

New association map:

$$
\begin{array}{lccccccccccc}
A_t: & \dots & \boxed{\mathbf{a1}} & \boxed{6} & \boxed{\mathbf{b}} & \boxed{9} & \boxed{\mathbf{a2}} & \boxed{8} & \boxed{\mathbf{a}} & \boxed{2} & \dots \\
B_t: & \dots & \boxed{\mathbf{a1}} & \boxed{6} & \boxed{\mathbf{a2}} & \boxed{9} & \boxed{\mathbf{b}} & \boxed{8} & \boxed{\mathbf{a}} & \boxed{2} & \dots \\
 & & \uparrow & & \uparrow & & \uparrow & & \uparrow & & \\
 & & d_1^*/d_2^* & & d_2^*/i_2 & & i_2/d_1^* & & i_1^* & &
\end{array}
$$

where at t_2,

$$
\boxed{\begin{array}{l}
d_1^*/d_2^* \longrightarrow i_2 \\
i_2/d_1^* \longrightarrow d_1^* \\
d_2^*/i_2 \longrightarrow d_2^*
\end{array}}.
$$

Chain of association maps.

$\boxed{\textbf{a2}}$ will do label-to-location jump w.r.t. the following assoc. map

$$
\begin{array}{llcccccccccc}
A_t: & \ldots & \boxed{\textbf{a1}} & \boxed{6} & \boxed{\textbf{b}} & \boxed{9} & \boxed{\textbf{a2}} & \boxed{8} & \boxed{\textbf{a}} & \boxed{2} & \ldots \\
B_t: & \ldots & \boxed{\textbf{a1}} & \boxed{6} & \boxed{\textbf{a2}} & \boxed{9} & \boxed{\textbf{b}} & \boxed{8} & \boxed{\textbf{a}} & \boxed{2} & \ldots \\
 & & \uparrow & & \uparrow & & \uparrow & & \uparrow & & \\
 & & i_2^{**} & & d_2^{**} & & d_1^{**} & & i_1^{*} & &
\end{array}
$$

d_1^{**} is i_2/d_1^* before t_2, and d_1^* after t_2;
d_2^{**} is d_2^*/i_2 before t_2, and d_2^* after t_2;
i_2^{**} is d_1^*/d_2^* before t_2, and i_2 after t_2.
$\ll \boxed{\textbf{a2}}, i_3 \gg$ occurs at $t_3 \sim$ **exponential** $\left((1-2/n)\cdot\frac{1}{n}\right)$

Chain of association maps.

And so on, creating a **chain** of $k = \lfloor \varepsilon n \rfloor$ association maps.

In case of $d = 2$ discrepancies, the average time of discrepancy cancelation on one of association maps is

$$E[T_2] = \frac{n^2}{4(k+1)} \approx \frac{n}{4\varepsilon}.$$

Result.

General d:

$$E[T_d] = \frac{n^2}{2(k+1)d} \approx \frac{n}{2\varepsilon d}.$$

Coupling time (all discrepancies):

$$E[T_{coupling}] \leq \left[\frac{1}{2\varepsilon} + \frac{\kappa}{(1-\kappa)(\kappa-\varepsilon)}\right] \cdot n \log n$$

for any $0 < \varepsilon < \kappa < 1.$